International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0278
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1759 - 1764

# Building A Symptom-Based Disease Diagnosis Web App with Flask and Machine Learning

*M. Akshith Reddy[1], V. Sai Manish Reddy[2], K. Vamshi Goud[3], E. Nithin Kumar[4], Mr. Kumar Baradur[5], Dr. M. Ramesh[6]*

*[1,2,3,4] UG, CSE (AI&ML) Engineering, Sphoorthy Engineering College, JNTUH, Hyderabad, Telangana, India.*

*[5]Assistant Professor, Department of Computer Science & Engineering (AI&ML), Sphoorthy Engineering College, JNTUH, Hyderabad, Telangana, India.*

*[6]Professor & Head of the Department, Department of Computer Science & Engineering (AI&ML), Sphoorthy Engineering College, JNTUH, Hyderabad, Telangana, India.*

***Emails**: macha.akshith@gmail.com[1], saimanishreddy55@gmail.com[2], vamshimaha182003@gmail.com[3], nithine57@gmail.com[4], kumar.baradur@gmail.com[5], hodaiml@sphoorthyengg.ac.in[6]*

## Abstract

*Building a symptom-based disease diagnosis web application built using Flask and machine learning, designed to assist users in identifying potential health conditions based on reported symptoms. The system leverages a trained machine learning model to analyze symptom data and predict possible diseases, providing a preliminary diagnosis and guiding users toward seeking appropriate medical advice. The algorithms used in various prediction system consisted of Linear Regression, Decision Tree, Naïve Bayes, KNN, Random Forest Tree,etc. by using these it is possible to predict more than one disease at a time. So, the user does not need to traverse many models to predict the diseases. The model employs algorithms optimized for multi-class classification, capable of handling complex symptom-disease relationships to improve diagnostic precision. Flask, a lightweight yet powerful web framework, serves as the application's backbone, providing a responsive interface that facilitates symptom input, rapid data processing, and real-time display of diagnosis results, ensuring a smooth user experience. Beyond basic diagnosis, the application offers a range of functionalities aimed at enhancing user engagement and education, including detailed information about description, precaution, medication, workout and diet related to that disease.*

***Keywords:** Machine learning, Linear Regression, Decision Tree algorithm, Naïve Bayes algorithm, KNN algorithm, Random Forest Tree algorithm, symptom based disease diagnosis.*

## 1. Introduction

Healthy lifestyle, healthcare and medicines are few of the essential elements of human lifestyles and economy. There is a tremendous change in the world we are living in now and the world that existed few months back [1]. Everything has turned ugly and divergent. In this case, where the entirety has grown to become digital or let us say virtual, the doctors and nurses are giving their maximum efforts to keep people's lives and people's health even though they ought to danger their very own. Even now in some parts of the world there are still some far-flung villages, remote places which lack clinical centres, health facilities. Machines have started to gain popularity and dependency by humans as, without any human mistakes, they could perform duties greater efficaciously and with a steady degree of accuracy. A disease predictor is nothing but a virtual doctor, which can predict the disorder of any affected person without any human errors [2]. There is a tremendous change in the world we are living in now and the world that existed few months back. Everything has turned ugly and divergent. In this case, where the entirety has grown to become digital or let us say virtual, the doctors and nurses are giving their maximum efforts to keep people's lives and people's health even though they ought to danger their very own. Even now in some parts of the world there are still some far-flung villages, remote places

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0278
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1759 - 1764

which lack clinical centres, health facilities. Machines have started to gain popularity and dependency by humans as, without any human mistakes, they could perform duties greater efficaciously and with a steady degree of accuracy. A disease predictor is nothing but a virtual doctor, which can predict the disorder of any affected person without any human errors [3]. There is a tremendous change in the world we are living in now and the world that existed few months back. Everything has turned ugly and divergent. In this case, where the entirety has grown to become digital or let us say virtual, the doctors and nurses are giving their maximum efforts to keep people's lives and people's health even though they ought to danger their very own. Even now in some parts of the world there are still some far-flung villages, remote places which lack clinical centres, health facilities. Machines have started to gain popularity and dependency by humans as, without any human mistakes, they could perform duties greater efficaciously and with a steady degree of accuracy [4]. A disease predictor is nothing but a virtual doctor, which can predict the disorder of any affected person without any human errors. This project revolves around the creation of an advanced web application aimed at diagnosing diseases based on user-inputted symptoms. Utilizing the Flask framework for its development, the application will seamlessly integrate machine learning algorithms to predict possible diseases. Users will be able to input their symptoms through a straightforward interface, and the machine learning model, trained on a comprehensive dataset, will analyze these inputs to provide a list of potential diseases. Each disease prediction will be accompanied by detailed information, including descriptions, precautionary measures, recommended medications, and suggested workout and diet plans to help manage the condition effectively [5].

## 2. Data Source and Statement

The healthcare sector faces significant challenges in providing quick and accurate disease diagnoses, especially in resource-limited settings where access to medical professionals and facilities is constrained. The traditional approach to diagnosing diseases based on symptoms often requires consultations with healthcare professionals, which can be time-consuming and inaccessible to many Symptom-Based Disease Prediction Model does not focus on the prediction of a specific disease; instead, it predicts disease based on the symptoms given by the user. Individuals [6]. A Symptom-Based Disease Prediction Model does not focus on the prediction of a specific disease; instead, it predicts disease based on the symptoms given by the user. As a result, the user does not need to traverse many models to predict the disease [7]. There is a probability of lowering the death rate due to the prediction of disease at an early stage. Healthcare systems worldwide face challenges in providing timely, accurate, and comprehensive disease diagnosis, particularly in regions with limited access to medical professionals [8]. While various diagnostic tools exist, there is a growing need for an intelligent system that not only diagnoses conditions but also provides holistic management recommendations (Table 1).

**Table 1 Some Rows of Disease with Them Corresponding Symptoms in The Dataset**

| | Disease | Symptoms |
|---|---|---|
| 1 | Malaria | {chills, vomiting, high_fever, sweating, heada... |
| 2 | Allergy | {continuous_sneezing, shivering, chills, water... |
| 3 | Fungal infection | {skin_rash, nodal_skin_eruptions, dishromic_... |
| 4 | Gastroenteritis | {vomiting,sunken_eyes, dehydration, diarrhoea |
| 5 | arthritis | {muscle_weekness, stiff_neck, swelling_joints,... |
| 6 | Typhoid | {chills,vomiting, fatirgue, high_fever, headac... |
| 7 | Hypertension | {muscle_weakness, stiff_neck, swelling_joint,.... |

The dataset used for machine learning requires features such as data contains 40 diseases and 133 symptoms which is around 5000 rows. The data is split in two parts in which 3400 rows are used for model training whereas 1500 is used for validation [9]. And last column for the disease class (40 unique

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0278
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1759 - 1764

disease classes). Some rows of disease with their corresponding symptoms in the dataset [10].

## 3. Proposed System and Methodology

### 3.1. Pre Processing

Effective data preprocessing is crucial. Techniques employed include removing noise, handling missing information, modifying default values, and potentially grouping attributes for prediction. The first column contains the disease and the subsequent columns consist of all the symptoms. -After collecting that data, as that data is raw data we have to make it suitable for training our machine learning model. By using some python libraries like NumPy, and pandas, we have made that data suitable for machine learning models.

### 3.2. Model Building

After applying these algorithms, we have to select which is most fitted with our dataset and which gives us more accuracy. So, we have used a confusion matrix for that and mapped out the accuracy of each model. And,All the symptoms are converted to a vector of 0 and 1 and out of 40 diseases one disease will have 1 in its index in the output vector whereas all the others will have 0. The disease which has the highest probability or 1 in its index is given as output. This is a machine learning classification problem. Here we would be using classification machine learning models like KNN, SVM, decision tree, random forest classifier, Naïve Bayes.

- P(y_pred) >= 0.7: Top 1 Disease
- P(y_pred) >= 0.5 and P(y_pred) < 0.7: Top 2 Disease
- P(y_pred) < 0.5: Top 3 Disease

**1) Support Vector Machine (SVM)**

In this algorithm, the data is placed in a multidimensional space and this data is classified into classes by generating a separation hyperplane between them. The main aim of this algorithm is to increase the distance between marginal lines that are there on either side of the hyperplane. Based on which side of the hyperplane the point lies, it is classified into classes.

**2)K Nearest Neighbor (KNN)**

This algorithm firstly finds K neighbors to the provided input based on distance and then assigns the majority class as the output. The K is a hyperparameter. for various values of K the accuracy is tested and the K value having best accuracy is taken.

**3)Decision Tree**

The decision tree algorithm works on the concept of entropy. The entropy is the measure of randomness i.e., if in the data there is 50% of class A and 50% of class B it means the entropy is 1 as the data is totally random but if the data contains 100% class A, then the entropy is 0. The decision tree splits the data in such a way that the entropy should finally become 0 and we are able to classify the classes clearly. So, at every step it selects a column splitting which could get least entropy; it continues the process until it gets 0 entropy.

**4)Random Forest Classifier**

Random forest is a bagging algorithm. It uses N number of decision trees, each decision tree gives its output and based on the majority vote of the class among all decision trees the output is decided.

**5.Naïve Bayes**

This algorithm uses Bayes Probability theorem to give the output. The Bayes theorem states that the probability of an outcome is dependent on the conditions that might be related to the outcome. Machine Learning models for symptom-based disease diagnosis (Refer Figures 1 to 5).

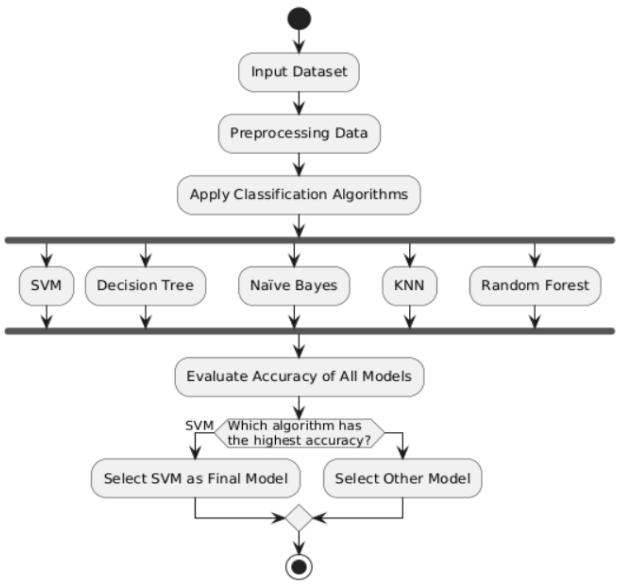


**Figure 1 System Architecture of Symptom-Based Disease to Personalized Recommendation Delivery**

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0278
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1759 - 1764

### 3.3. Web Interface Deployment

The technical infrastructure of the application is designed for scalability and accessibility. Flask, a lightweight and efficient web framework, handles the backend, including routing user requests and serving the machine learning model's predictions. A robust database—implemented using SQLite—stores disease information, while libraries like scikit-learn, pandas, and NumPy power the machine learning pipeline. The frontend, built with HTML, CSS, and JavaScript, ensures that the interface is intuitive and responsive, making it accessible across devices.

**Key Features of the Interface Include:**

- **Precautions:** Provides infection prevention tips, lifestyle modifications, stress management, and Real-Time Alerts for high-risk patients, warning them about environmental triggers.
- **Medication:** Includes Dosage Optimization based on personal metabolism and health history, and Drug Interaction Warnings to ensure safe combinations of medications.
- **Workout Plan:** For patients recovering from injuries or with limited mobility, the system can suggest safe exercises or physical therapy regimens.
- **Diet Plan:** By integrating nutritional data, genetic factors, and disease-related requirements, the system provides a customized diet plan.
-

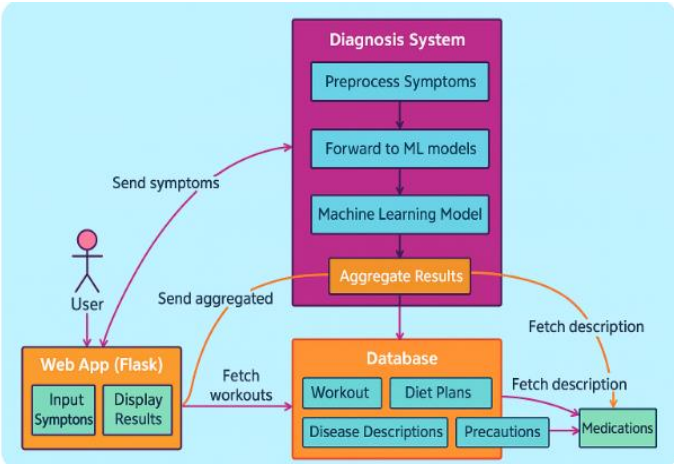


**Figure 2 System Architecture of Symptom-Based Disease from Data Ingestion to Personalized Recommendation Delivery**

## 4. System Testing and Results

### 4.1. System Testing

To ensure the reliability, stability, and correctness of the platform, comprehensive testing was conducted across all components and modules. The testing process incorporated multiple types of evaluation, including unit testing, where individual components such as symptom input validation, machine learning model predictions (SVM, KNN, Naïve Bayes, Decision Tree, and Random Forest), and database queries are verified for accuracy. Once unit testing is complete, integration testing is conducted to check the communication between Flask routes, the machine learning prediction engine, and the database, ensuring smooth data flow. Functional testing further validates that the app correctly predicts diseases, retrieves detailed descriptions, and provides precautionary measures, recommended medications, workout routines, and diet plans, ensuring users receive comprehensive health guidance.

```
In [18]:
# selecting svc
svc = SVC(kernel='linear')
svc.fit(X_train,y_train)
ypred = svc.predict(X_test)
accuracy_score(y_test,ypred)

Out[18]: 1.0
```

**Figure 3 Selecting SVC Model**

Beyond core functionality, performance testing is carried out to measure response time and scalability, ensuring that the web app can handle high traffic efficiently. Security is another critical aspect, so security testing is done to validate user authentication, encryption of sensitive data, and protection against vulnerabilities like SQL injection and cross-site scripting. Usability testing ensures the app's interface is intuitive and accessible, providing a seamless experience for users of different

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0278
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1759 - 1764

backgrounds and technical expertise. These test phases collectively ensure that the system functions reliably, efficiently, and securely.

```
# test 1:
print("predicted disease :",svc.predict(X_test.iloc[0].values.reshape(1,-1)))
print("Actual Disease :", y_test[0])
```

```
predicted disease : [40]
Actual Disease : 40
```

**Figure 4 Evaluation**

### 4.2. Results

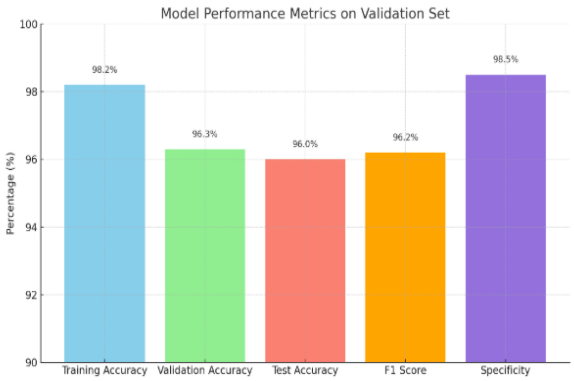


**Figure 5 Results**

In this project, a symptom-based disease diagnosis web application was developed using Flask and Machine Learning, leveraging a dataset containing 133 symptoms across 40 different diseases. After tuning and comparing the models, SVM with an RBF kernel was selected due to its superior performance across multiple evaluation metrics. The SVM model achieved a training accuracy of 98.2%, a validation accuracy of 96.3%, and a test accuracy of 96.0%, demonstrating excellent generalization. In terms of classification metrics, the model reached a macro-averaged F1 Score of 0.962 and an average specificity of 0.985, indicating strong performance across all 40 disease classes. The web application allows users to select from a checklist of symptoms and, upon submission, predicts the most likely disease. Alongside the prediction, the app displays a comprehensive paragraph containing a description of the disease, precautionary measures to be taken, commonly recommended medications, and tailored workout and diet plans for recovery and health maintenance. Continuous model refinement and dataset updates will further enhance prediction accuracy and user experience, making the web app an effective tool for disease diagnosis and health management.

## Conclusion

This paper presented a comprehensive overview of the symptom-based disease diagnosis web app that leverages Flask and machine learning algorithms (SVM, KNN, Naïve Bayes, Decision Tree, and Random Forest) to provide users with personalized health information. By inputting their symptoms, users receive detailed disease descriptions, precautionary measures, recommended medications, and tailored workout and diet plans. This app aims to enhance healthcare accessibility, empower users with valuable health insights, and support early detection and management of diseases. With a user-friendly interface, adherence to data privacy standards, and real-time predictions, the app bridges the gap between users and healthcare professionals, offering a valuable tool for health management and education. By continuously improving and integrating user feedback, the app ensures reliability, scalability, and effectiveness in delivering personalized health solutions. Looking forward, several future enhancements can be considered to further strengthen the platform's capabilities. To enhance the symptom-based disease diagnosis web app, several features can be integrated. First, the app can be connected with wearable devices like fitness trackers and smartwatches to automatically collect health data such as heart rate, sleep patterns, and physical activity. This integration provides more comprehensive health insights and enables real-time monitoring of health metrics. Additionally, implementing a chatbot to guide users through symptom input can make the process more interactive and user-friendly. Generating personalized health reports that users can download and share with their

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0278
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1759 - 1764

healthcare providers will facilitate better communication and informed medical consultations. Implementing a feedback mechanism allows users to report issues and suggest improvements, ensuring continuous enhancement of the app.

## Acknowledgements

We express our sincere gratitude to the medical professionals and data scientists who contributed to the collection and verification of clinical datasets. We also acknowledge the use of Python-based tools and libraries such as Scikit-learn, Pandas, and Matplotlib in developing and validating this model. Lastly, appreciation is extended to the healthcare institutions and mentors for their guidance and domain expertise.

## References

[1]. Grampurohit S and Sagarnal C, "Disease Prediction using Machine Learning Algorithms," In 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154130

[2]. Singh A and Kumar R, "Heart Disease Prediction Using Machine Learning Algorithms," In 2020 International Conference on Electrical and Electronics Engineering (ICE3), 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958

[3]. Dahiwade D, Patle G and Meshram E, "Designing Disease Prediction Model Using Machine Learning Approach,"In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 1211- 1215, doi 10.1109/ICCMC.2019.8819782

[4]. Hamsagayathri P and Vigneshwaran S, "Symptoms Based Disease Prediction Using Machine Learning Techniques," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 747-752, doi: 10.1109/ ICICV50876. 2021. 9388603.

[5]. Xu X, Huang X, Ma J and Luo X, "Prediction of Diabetes with its Symptoms Based on Machine Learning," 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE), 2021, pp. 147-156, doi: 10.1109/CSAIEE54046.2021.9543343.

[6]. Kohli P S and Arora S, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777449.

[7]. Sharma V, Yadav S and Gupta M, "Heart Disease Prediction using Machine Learning Techniques," 2020 2ndInternational Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 177-181, doi: 10.1109/ICACCCN51052.2020.9362842.

[8]. Srivastava, A. K. (2023). Review Study of Contemporary Work in Crop Yield Prediction Using Machine Learning Models.

[9]. Singh, V.; Asari, V.K.; Rajasekaran, R. A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease.

[10]. Priyanka Rastogi, Kavita Khanna, Vijendra Singh, LeuFeatx: Deep learning–based feature extractor for the diagnosis of acute leukemia from microscopic images of peripheral blood smear, Computers in Biology and Medicine, Volume 142,2022